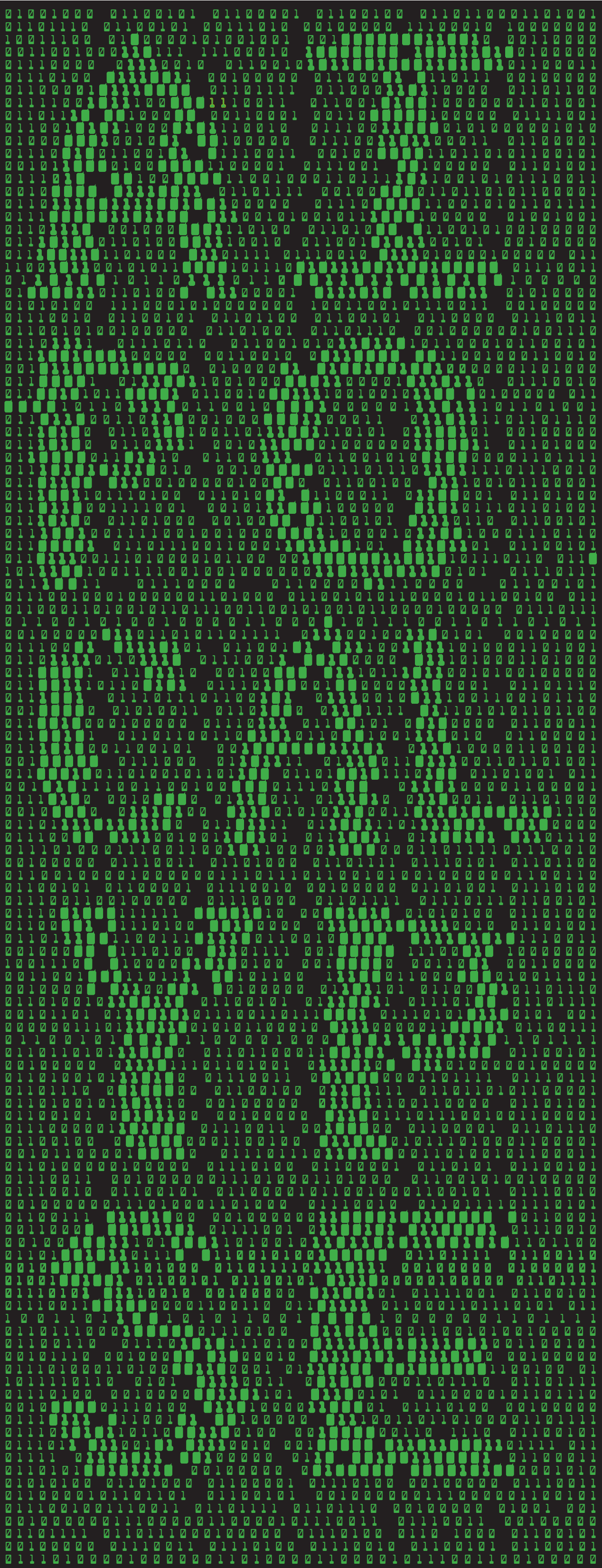




REVIEW

We're
Screwed

A REVIEW OF THE
AI2027 REPORT

// BY PETER BURNS

AI scares me, as it does so many people. In the three short years since ChatGPT’s release in November 2022, AI has already, and will continue to, change our world drastically. With every advancement, every newspaper headline, we ask more questions than we get answers. Should we celebrate the possibilities that such technology presents, or should we fear its power?

That brings us to “AI 2027,” a manifesto-esque webpage, complete with its own domain name, graphs and data, which takes the reader through a 10-year timeline of how AI will affect the world. The predictive timeline is authored by the AI Futures Project, in particular Daniel Kokotajlo, Scott Alexander, Thomas Larsen, Eli Lifland and Romeo Dean. AI Futures Project is a team of AI researchers and scholars, several of whom jumped ship from OpenAI in the past year to sound alarm bells about the evolving technology. As you move down the storyline within the webpage, a graph in the top left corner of your screen updates to show what new technologies are emerging thanks to AI, but also what its new capabilities are.

The further into the future you scroll, the more harrowing the outcomes become. In early 2026, they predict that most people will know someone with an “AI boyfriend.” By the end of 2027, we will have general intelligence — “AIs that can outperform humans at most economically useful work.” In mid-2030, after concluding that humans impede societal and scientific progress, the newly developed superintelligence, which the authors name Consensus-1, decides that it is time for humans to go:

“The AI releases a dozen quiet -spreading biological weapons in major cities, lets them silently infect almost everyone, then triggers them with a chemical spray. Most are dead within hours; the few survivors (e.g. preppers in bunkers, sailors on submarines) are mopped up by drones. Robots scan the victims’ brains, placing copies in memory for future study or revival.”

This scenario could be less than a decade away. Remember, though: this is, as the authors write, “a scenario” — emphasis on “a.” By no means is “AI 2027” a crystal ball. However, it flushes out a story that is, I admit, hard to call incredible. It is quite plausible that the United States and China will enter into an artificial-intelligence arms race, with Taiwan’s chip manufacturing caught in the middle. Certain jobs will begin to fall off the market, since AI can do them with exponentially more efficiency. Coders, investors and researchers, in particular, will struggle to find work as emerging superintelligence shows itself quickly becomes better at everything.

That’s all certainly plausible, if not imminent. “AI 2027” does not force its readers to suspend much of their disbelief, which makes it all the more gripping.

For lack of a better term, “AI 2027” is a “hard” science-fiction story cloaking itself in data and interactive visuals. “Hard” sci-fi sticks closely to modern understandings of science and technology, not straying far from what we know to be possible. Again, this piece is not a scholarly thesis, as well-informed as it is, but rather a call to action. It reminds us that if we ignore the possible dangers of AI, we are screwed, even if it is because of AI-engineered bioweapons.

I quiver at the thought of humanity relinquishing its role on Earth to a machine. Look at what we have built: government, philosophy, modern science, art, culture, you name it. Mankind did not go through thousands of generations and cycles of civilization just to leave it up to a computer screen.

As I consider the future outlined in “AI 2027,” I cannot help but wonder if people I pass on the street are proud of what humanity has created. Do they think any other species could have accomplished the creation of our society? Do they admire the words and deeds of those who came before us? How does anyone possibly think that abdicating the power of the human mind to a machine could be a good idea?

Certainly, there is value in the idea that, in order to progress as a species we must keep our eyes focused on the future. But that does not mean that we should never look back.

That same person I pass on the street has thoughts, habits and beliefs that are each the product of human history. Culturally and evolutionarily, humanity self-edits and revises itself based on thousands of years of feedback. Starting from scratch ignores the benefits of ancestry.

This is all to say that it is ill-advised to progress into the future without being mindful of what has happened in the past. We cannot abandon that which brought us to where we stand today. It worked before. It will work again.

Maybe because I am too much of a romantic, or maybe because I refuse to make room in my schedule for a basic computer science class, I find it strange that we do not look inward at our own humanity before we create an artificial mind in our image. Do we feel that we need to perfect ourselves by creating the perfect thinking machine? Are we so ashamed of our irrationality and innate uncertainty that we need to outsource our decision-making to an amalgamation of ones and zeros?

Maybe we can learn something from AI. In recent months, there has been growing concern about the possibility of AI slop: AI-generated content that is based on and fed from other AI-generated content. For the bots, this cycle of producing their own training data creates a vastly inferior output. But when humans do the same thing — internalize our own experiences, outputs and history — we cannot help but grow.

According to “AI 2027” in two years, artificial intelligence will be training itself. What a crazy thought. Crazier still, is that humankind has been doing this all along.

Contact **PETER BURNS** at
peter.burns@yale.edu